ENHANCED BREAST CANCER IDENTIFICATION AND IMPACT ANALYSIS VIA MODIFIED FUZZY C-MEANS INTEGRATED WITH SOFT COMPUTING TECHNIQUES

A SYNOPSIS

Submitted by

S.BALAJI

(Reg. No. 201608238)

in partial fulfilment for the award of the degree

of

DOCTOR OF PHILOSOPHY

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

SCHOOL OF ELECTRONICS, ELECTRICAL AND BIOMEDICAL TECHNOLOGY

KALASALINGAM ACADEMY OF RESEARCH AND EDUCATION

(DEEMED TO BE UNIVERSITY)

ANAND NAGAR, KRISHNANKOIL – 626 126

TAMILNADU, INDIA.

DECEMBER 2023

1. INTRODUCTION

Breast cancer is one of the prevalent diseases with rising mortality rates among women, and these rates continue to rise annually. The only way to control breast cancer mortality rates is through early detection and periodic screenings. However, a dearth of awareness causes individuals to disregard this healthy practice, resulting in negative outcomes. Due to advancements in medical imaging science, automated medical image analytic systems are attracting the interest of researchers. Mammography images are the most common and histopathological images provide the clearest view of tissue structures among all imaging modalities, as their fundamental structure is preserved to the greatest extent.

Typically, histopathological images are used to determine the presence of disease, the disease's severity, and the disease's adverse consequences. Nonetheless, analyzing and presenting an automated analytical system based on histopathological images represents a significant challenge. Histopathological images are stained and may appear irregular and in some instances, the image's components appear imprecise and overlapping, which makes analysis even more difficult. Understanding these challenges, this research study intends to propose three different automated systems for breast cancer detection system for mammography and histopathological images.

The initial phase of research presents a 'CADS' for breast cancer detection and classification using mammogram images. The entire work is divided into four main phases, each of which plays an important role in achieving a better outcome for the proposed work. Lifting Wavelet Transform (LWT) is used for pre-processing, and the Lion Optimized Algorithm (LOA) based Fuzzy C Means (FCM) algorithm is used for segmentation. The Gabor and Local Ternary Pattern (LTP) features are extracted from the segmented images and the classification is conducted by Support Vector Machine (SVM).

The next of research proposes a mitotic cell detection and classification system based on supervised learning and the four significant image processing processes of pre-processing, area of interest extraction, potential feature extraction, and classification. In the pre-processing phase, stain normalization, denoising, and image smoothing are conducted, and nuclei are extracted using the Whale

Optimization Algorithm (WOA) based Generalized Hierarchical Fuzzy C Means (GHFCM) area of interest process. From the segmented regions, potential features are extracted, and the Kernelized Extreme Learning Machine (KELM) is used to classify mitotic cells.

The final phase of research presents an ensemble learning based mitotic cell classification. The potential of ensemble classification over singleton classifier is proven at the cost of time consumption, however is reasonable. The performances of all the proposed research solutions are evaluated in terms of standard performance measures and the efficiencies of the proposed solutions are proven with better accuracy, sensitivity and specificity rates at reasonable time consumption rates.

2. LITERATURE REVIEW

This section studies the recent literature concerning the breast cancer detection and diagnostic system.

Ker J et.al [1]presented an automated brain histology classification system with machine learning concept. This work considers both the brain and breast tissues with the help of Convolutional Neural Networks (CNN). The histology images are classified into normal, Low Grade Glioma (LGG) and High Grade Glioma (HGG).

Kausar Tet.al [2]presented a multi-class recognition system based on Haar wavelet based CNN meant for breast histopathology image.The wavelet applied images are then operated by CNN and multi-scale features are extracted by the CNN model. This work employs data augmentation and transfer learning approaches.

Xu J et.al [3]presented a deep CNN based system for the process of segmentation and classification of epithelial and stromal regions of histopathological images. The color and texture features are extracted along with high level features for performing classification.

Beevi K. S et.al [4] presented an automatic mitosis detection scheme for breast histopathological images. This work utilizes random forest classifier in association with CNN. The unique features are extracted from the nuclei patches and the CNN performs classification.

Tashk A et.al [5] discussed an automatic identification of mitotic cells from breast cancer images. This work relies on texture, statistical and mathematical features. A mathematical model based on statistical moment entropy and stiffness matrix is created. This work employs Support Vector Machine (SVM) as a classifier.

Öztürk Ş., & Akdemir B. [6]presented a classification scheme for deep CNN for histopathological images of breast. This work detects the cancerous regions available in the images and the classification is attained by window based technique.

Das D. K., & Dutta P. K.[7] presented an automated scheme for the detection of mitotic cells of breast cancer histological images on the basis of CNN and wavelet.

Öztürk Ş., & Akdemir B.[8] discussed a phase classification method of mitotic events with the help of dictionary learning for stem cell populations. This work creates a dictionary, which represents the phases in single feature space and this work creates three dictionaries with varied features. Additionally, a selective dictionary based on the automatic choice of the optimal values is presented.

Kaushal C et.al [9] reviewed a current trends in breast cancer diagnosis with histopathological images. This work considers the techniques being proposed in the last ten years.

Tandon Aet.al [10] presented a study on image analysis for mitosis detection with different stains. This work considers mitotic images with respect to oral mucosa, dysplasia and squamous cell carcinoma, which are stained by Hematoxylin and Eosin (H and E), crystal violet.

Bolhasani Het.al [11] presented a histopathological image dataset for the purpose of grading breast invasive ductal carcinomas. In this paper, an image dataset with 922 samples is presented with fifty specimens of each grade.

Roy K et.al [12] presented a patch based system for classifying the breast histology images with the help of deep learning. This work presents two modes of classification, which declares a result for every single patch and a single result for all the patches available.

Gandomkar Z et.al [13]proposed a work to classify breast histopathological images with deep residual networks. This work considers the benign form of cancer

and further classifies it into ductal, lobular, mucinous and papillary carcinoma. A meta-decision tree is built for achieving classification.

Roullier Vet.al [14] presented a multi-resolution enabled graphical analysis of histopathological images for extracting mitotic cells and visualizing them. The images are analysed in different resolution levels and the spatial refinement is carried out by means of label regularization.

Maroof N et.al [15]presented a mitosis detection scheme for histopathological images of breast cancer with the help of hybrid feature space. This work aims to discriminate between mitotic and non-mitotic nuclei with the help of color and texture features. The SVM classifier is employed for carrying out classification.

Wahab Net.al [16] presented a two-phase deep CNN based breast cancer detection system. Initially, the nuclei are extracted and the CNN is trained with various patches of pixels. The phase-I classification is again modified by the phase-II for reducing the imbalance.

3. OBJECTIVES

The central theme of this research is to present automated breast cancer detection system, where the initial research phase presents a breast cancer classification system, while the remaining research phases tend to classify between the mitotic cells for measuring the severity of cancer.

I. To present a breast cancer detection system based on optimized segmentation scheme and supervised learning algorithm.
II. To propose supervised learning scheme based mitotic cell detection and classification system on histopathological images of breast cancer.To present an ensemble classification based mitotic cell classification scheme on histopathological images of breast cancer.
III. To reduce false positive and false negative rates of the proposed mitotic cell classification systems.
IV. To develop a breast cancer detection system with the combination of PSO (Particle Swarm Optimization) Spatially Constrained Adaptively Regularized Kernel based FCM (ScARKFCM).

The overall objectives of this research work are listed above and the following section outlines the research overview.

4. RESEARCH PROBLEM

The increased human mortality rate attributable to breast cancer is the primary impetus for conducting this study. According to the report, approximately 15% of cancer-related deaths are attributable to breast cancer. It is well-known that breast cancer is prevalent among women, and the report indicates that approximately 15% of cancer-related deaths are attributable to breast cancer. Histopathological images play a vital role in studying and analyzing the severity of the disease, resulting in an accurate diagnosis.

Nonetheless, manual analysis of histopathological images is quite challenging, as it requires a significant amount of prior knowledge and is also a tedious process. In certain instances, the incorrect interpretation of histopathological images can directly impact the patient's life. All of these issues can be resolved by combining manual labor with machinery expertise.

The automated systems can assist the healthcare professional, but the healthcare professional can make the definitive diagnosis. This improves the diagnostic precision while consuming minimal time. This research presents three automated mitotic cell classification systems for histopathological images of breast cancer with higher accuracy rates for mitotic cell classification. The objectives of the research are as follows.

5. CONTRIBUTIONS

The main goal of this research is to present automated breast cancer detection and mitotic cell classification systems with better accuracy rates in addition to reasonable time consumption.

5.1 Computer Aided Diagnostic System for Breast Cancer Detection Based on Optimized Segmentation Scheme and Supervised Algorithm

Breast cancer is a serious threat to the womankind and it leads the susceptible kinds of cancer for women. The mortality rates due to breast cancer increases every

single year and the World Health Organization (WHO) aims to reduce the occurrence of breast cancer by at least 2.5% per year. The occurrence of breast cancer can be minimized only when periodical screening is carried out. Mammography is one of the effective screening procedure, which can effectively locate earlier signs of breast cancer.

As an aid, this work aims to present a system for the breast cancer detection and classification. This work is segregated into four phases and all these phases aim to enhance the classification performance. The efficiency of the proposed work is evaluated against the state-of-the-art approaches and the proposed CAD system proves 98.2% accuracy, with minimal false positive and false negative rates in a reasonable period of time. The contributions of this work are as follows.

- The LWT is employed to pre-process the mammograms, owing to its minimal time and memory requirements.
- The segmentation accuracy is improved, as it is performed by the combination of LOA and FCM algorithm. The optimal initial seed point is chosen by the LOA algorithm, which enhances the segmentation efficiency.
- The combination of Gabor and LTP features is effective in determining the significant features.
- SVM classifier requires minimal time for classification, as the training process is effectively done.
- The first phase of research is simulated in a stand-alone computer with 16 GB RAM using MATLAB R2017B upon the benchmark dataset MIAS. This dataset mainly involve three classes such as normal, malignant and benign. The MIAS dataset is composed of 326 images, which fall into three categories such as fatty, fatty and dense glandular. All the images are eight bit gray scale images of size 1024×1024. This work utilizes sixty percent of the images for training and the remaining images for testing. Figure 1. Shows the segmentation result of FCM-LOA and some of the sample results attained by the proposed work is shown in Table 1.

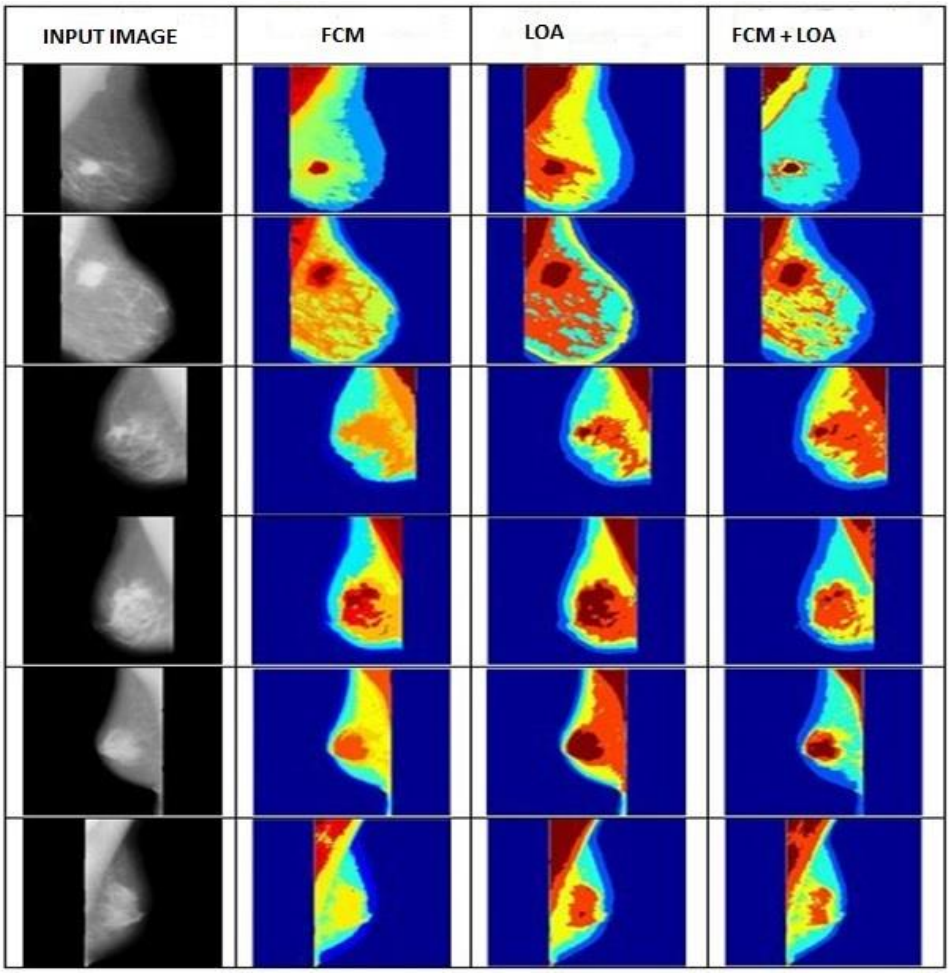


Figure 1: Samples of segmented output results.

Table 1: Performance analysis of Research Phase I

Techniques/Perf.measures	Accuracy (%)	Sensitivity (%)	Specificity (%)	Execution Time (s)
FCM	81.7	79.4	73.4	4.2
LOA	86.4	84.3	81.6	3.9
FCM+LOA	**98.2**	**98.6**	**96.7**	**2.7**
Gabor	81.4	76.34	71.8	4.8
LTP	88.3	83.9	79.7	4.6
Gabor+LTP	**98.2**	**97.9**	**96.6**	**2.7**
k-NN	74.3	69.4	71.8	11.62
RVM	82.7	79.8	79.7	4.2
SVM	**98.2**	**98**	**96.6**	**2.7**
MBO [6]	91.6	88.9	89.2	3.9
ML with LPP [12]	93.8	90.2	88.6	3.4
Global features [13]	96.1	91.3	91.4	3.1
Proposed	**98.2**	**98**	**96.6**	**2.7**

5.2 Expeditious and Precise breast cancer identification and classification using PSO based Spatially Constrained ARKFCM on clinical applications

The leading cause of death from cancer among women is breast cancer, which poses a major threat to their health. China is currently experiencing the highest mortality rate of breast cancer in the entire globe, and it is increasing. Due to the tremendous harm that breast cancer causes to one's health and well-being, research groups and organizations are working nonstop to develop the ideal diagnostic process and course of action. In certain sectors, artificially intelligent algorithms have been enabled to take the place of human behavior and judgement because of advancements in computer technology and learning algorithms. Medical professionals must frequently review patient data in order to diagnose breast cancer using the standard method.

In this instance, the algorithm is employed to swiftly provide clinicians with an extremely probable reference result, which is very crucial to improve the effectiveness of the diagnosis and lessen the strain on doctors. This research develops and implements a strategy based on a particle swarm optimization (PSO) algorithm to continuously alter the vital variables of the spatially constrained adaptively regularized kernel based fuzzy c means (ScARKFCM) clustering in order to increase the accuracy of cancer identification. The MIAS dataset was used to validate the developed PSO-ScARKFCM method's results. By locating and extracting the tumor from the raw medical images, the results aid clinicians in making an earlier diagnosis of the disease.

The dissection outcomes of the developed PSO-ScARKFCM on breast cancer images are portrayed in Figure 2. The performance metric parameters attained for the developed PSO-ScARKFCM technique for detecting tumor on breast images are ordered in Table 2.The MIAS dataset contains both the benign and malignant breast cancer images.

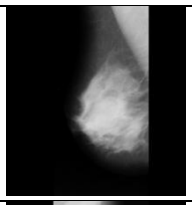
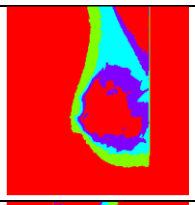
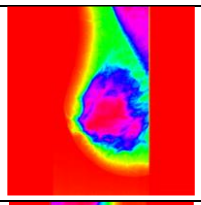
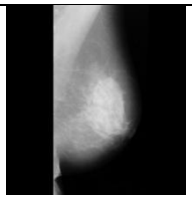
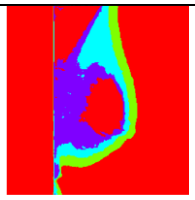
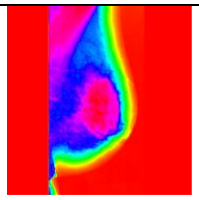
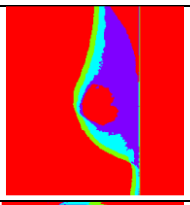
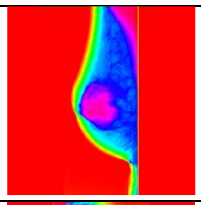
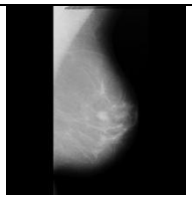
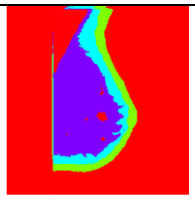
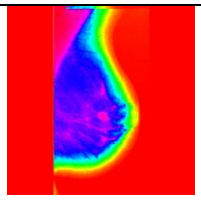
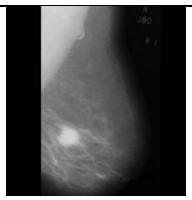
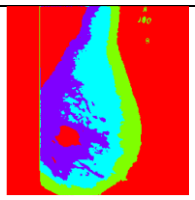
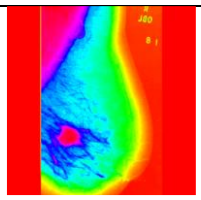
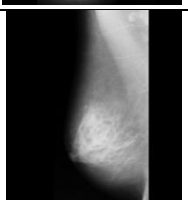
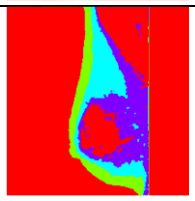
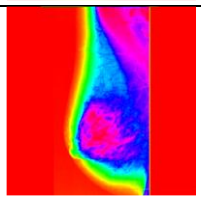
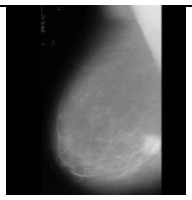
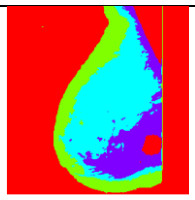
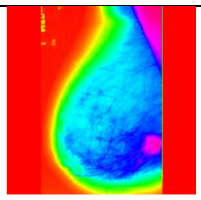
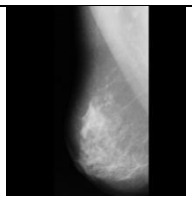
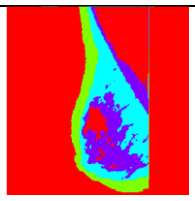
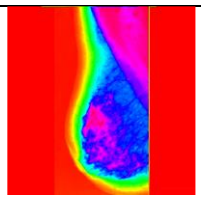
Image Type	Input Image	ScARKFCM	PSO-ScARKFCM
Benign			
Benign			
Benign			
Benign			
Malignant			
Malignant			
Malignant			
Malignant			

Figure 2: Segmented results using PSO-ScARKFCM method

Table 2: Performance evaluation on dissection breast cancer

PERFORMANCE METRICS	METHODOLOGIES		
	FCM	ScARKFCM	PSO-ScARKFCM
CT (in Seconds)	36.4	24.8	30.2
MSE	0.124	0.08	0.006
PSNR (in db)	58.2	62.4	65.5
TC (in %)	78.4	81.6	86.4
DS (in %)	76.2	82.1	89.4
ACCURACY (in %)	88.2	90.4	92.6
PRECISION (in %)	87.4	89.2	91.4
SENSITIVITY (in %)	89.4	91.8	92.2
SPECIFICITY (in %)	90.2	92.2	91.8

5.3 Supervised Learning Scheme based Mitotic Cells Detection and Classification in Breast Cancer Histopathological Images

Breast cancer shows the greatest mortality rates among women all over the world however, the mortality rates can be minimized when the disease is detected at its early stage. Hence, periodical screening is necessary such that the disease can be detected and treated on time. An automated breast cancer detection system paves way for effective diagnosis and reduces the false alarming rates. Histopathological images are quite difficult to process, owing to its intricate details and staining, yet they are effective.

Understanding the significance of mitotic cell detection, this paper proposes a supervised learning scheme based mitotic cell detection and classification, which is more reliable and promising. The mitotic and non-mitotic cells are classified by Kernelized Extreme Learning Machine (KELM) classifier, while the areas of interest are chosen by optimized clustering algorithm. The work performance of the proposed approach is tested by measures such as accuracy, sensitivity, specificity and time consumption. The performance of the proposed work is found to be satisfactory with minimized false positive and negative rates in contrast to the existing approaches. The important key points of this article are listed below.

- The combination of Generalized Hierarchical Fuzzy C Means (GHFCM) and Whale Optimization Algorithm (WOA) captures the area-of-interest better, which leads to better analytics.

- Employment of KELM classification scheme boosts up the reliability of the system, while reducing the FP and FN rates of classification.

- The effectiveness of the work is satisfactory with better sensitivity and specificity rates.

Accurate segmentation of the nuclei helps in achieving better classification results.This phase intends to detect and classify mitotic cells is analysed upon employing MATLAB 2017a version on a computer with 8 GB RAM. Mitos dataset is utilized for performance evaluation and it contains histopathological images with size $1376 \times 1539 \times 3$ and stained by H&E dyes.

The KELM classifier is imparted with knowledge for detecting the mitosis with the probability of 0.6. The train-test ration of the proposed work is set as 60-40 percent. The accuracy rates of the proposed work are ascertained by the ground truth image and the attained results are compared with the recent works such as Mathematical features [8], Haar based CNN [5], Deep learning [7], Neural Networks [9] and the results are shown in Table 3. and the comparison of performance measure of proposed workhas been shown in Figure 3.

Table 3: Performance comparison of Research Phase III

Performance Measures/ Classifiers	Mathematical Features [8]	Haar based CNN [5]	Deep Learning [7]	Neural Networks [9]	Proposed work
Accuracy (%)	89.7	91.4	94.2	97.2	**98.9**
Sensitivity (%)	86.2	89.8	91.1	93.7	**96.4**
Specificity (%)	83.4	86.2	89.4	92.9	**97.3**
F-measure (%)	84.77	87.96	90.24	93.29	**96.85**
Time consumption (s)	24.6	23.1	19.2	18.7	17.26

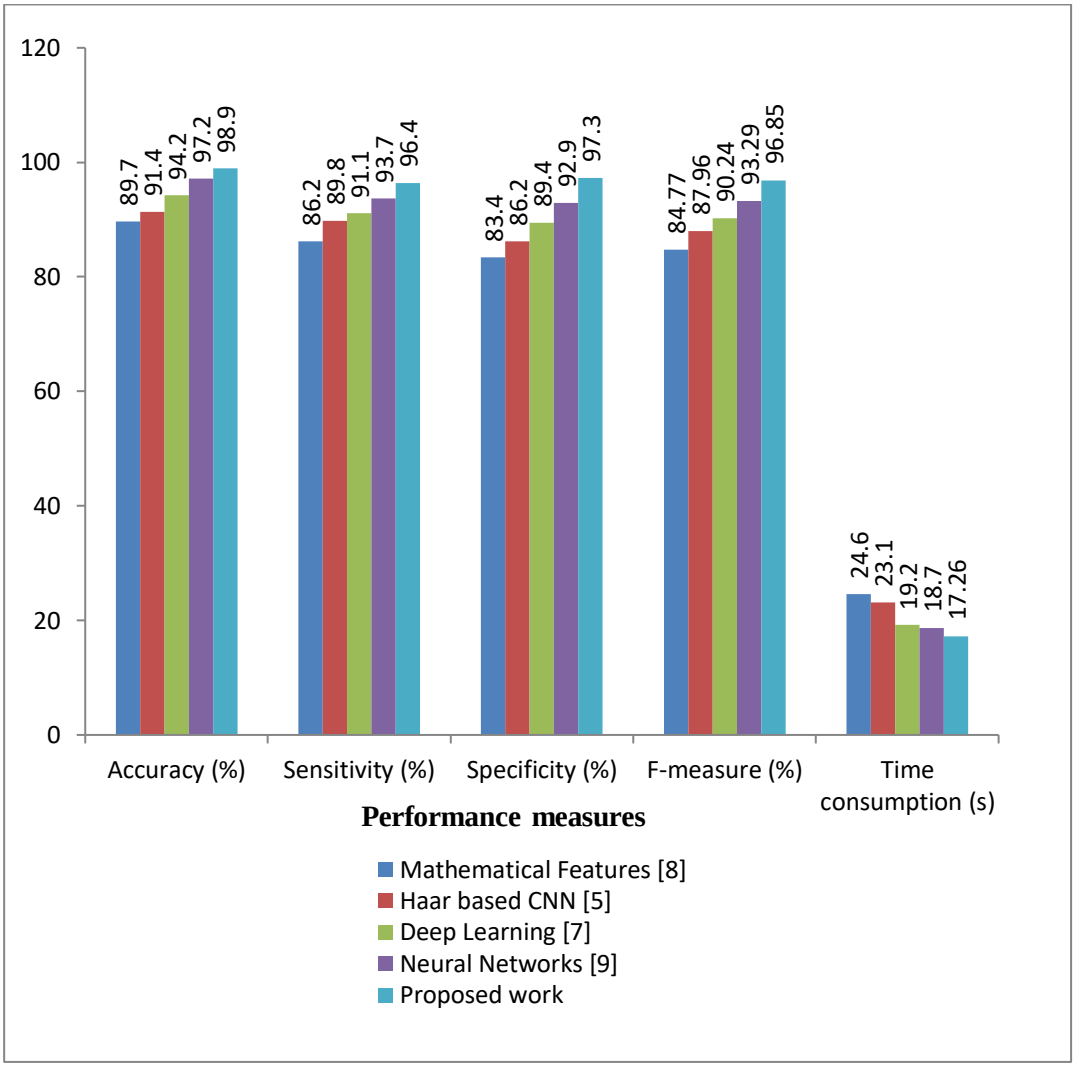


Figure 3: Comparison of the performance measure of proposed work

5.4 Ensemble Earning Based Mitotic Cell Classification System For Breast Cancer

On the list of life-threatening diseases that have greater death rates for the female population, breast cancer is at the top of the list. However, mortality rates caused by breast cancer can be reduced to a manageable level by the implementation of routine screenings. Images obtained by histopathology are consulted by pathologists for making a diagnosis or determining the stage of a malignant development. Because the histopathological images are so detailed, performing manual analysis on the images can be a challenging task.

This phase proposes an automated method for detecting mitotic cells and classifying them in breast cancer histopathology images. The performance of the suggested method is evaluated using common performance indicators such as accuracy, sensitivity, specificity, and time consumption in order to evaluate its effectiveness. The performance of the technique that was proposed is superior to the performance of the approaches that already exist. The key points of this phase are

- The histopathological images are treated with AMF and CLAHE for noise removal and contrast enhancement.
- KFCM and GWO are employed for segmenting nuclei and CLBP+SIFT features are extracted from the images.
- Ensemble classification approach is employed for decision making concerning the mitotic cell classification.

Hence, the overall views of all the research phases are discussed and the following section discusses the sample results attained by the proposed works.

The final phase of the work employs ensemble classification for mitosis detection. Based on the findings, it has been determined that the performance of the ensemble classifier is superior to that of the singleton classifiers such as Random Forest (RF), SVM, and ELM classifiersand the results are shown in Figure 4.

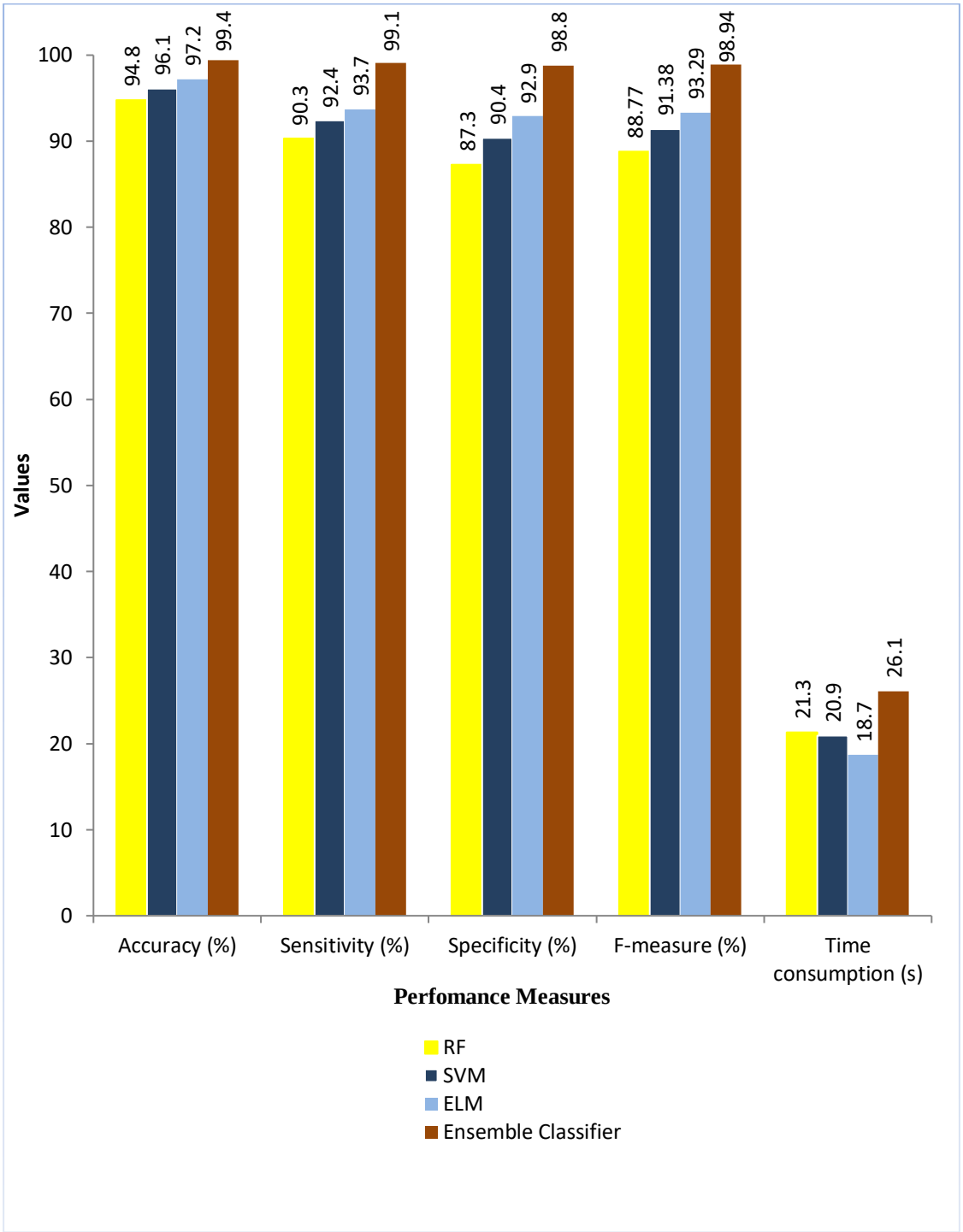


Figure 4: Performance analysis on ensemble classification

Hence, the potential of the proposed work is proven with the experimental results with better accuracy and F-measure rates at the rate of reasonable time consumption.

6. CONCLUSION

This research shows four different automated methods for detecting breast cancer and identifying mitotic cells from non-mitotic cells. The first two phase of research works on mammography images and the next two phase involves histopathological images. This study is broken up into four distinct phases, and the overarching goal of each of these phases is to improve the classification performance. The effectiveness of the suggested work is compared to the state-of-the-art methods and the proposed CADS is shown to have an accuracy of 98.2% with minimal rates of both false positive and false negative results in an acceptable amount of time. While considering the histopathological imagesthe Kernelized Extreme Learning Machine (KELM) classifier is used to differentiate between mitotic and non-mitotic cells, while the optimal clustering technique is responsible for selecting the areas of interest. In comparison to the existing methods, it has been determined that the performance of the work that has been proposed is adequate, with a reduction in the number of false positive and negative results. Hence the proposed supervised learning systems are more reliable and more promising for detection and classification of mitotic cell in multimodal images.

8. ORGANIZATION OF THE THESIS

The whole thesis is categorized into six chapters and the outline of each and every chapter is presented below.

This thesis is systematized into six chapters. Chapter 1 introduces the background of the research in addition to the research motivation, objectives and so on.

The related review of literature is discussed in chapter 2 and the basic concepts of digital image processing are also discussed in this chapter. Chapter 3 presents a CAD system for breast cancer detection, which is based on optimized segmentation scheme and supervised learning approach.

Chapter 4 discusses the mitotic cell classification system, which is based on supervised learning. Chapter 5 improvises the performance of the solution presented

in chapter 4, by presenting an ensemble classification based mitotic cell classification for breast cancer detection.

The thesis is concluded by chapter 6 with remarks, contributions in addition to future research directions.

REFERENCES

1. Ker J., Bai Y., Lee H. Y., Rao J., & Wang L. Automated brain histology classification using machine learning. *Journal of Clinical Neuroscience*, **66** (2019), 239-245.
2. Kausar T., Wang M., Idrees M., & Lu Y. HWDCNN: Multi-class recognition in breast histopathology with Haar wavelet decomposed image based convolution neural network. *Biocybernetics and Biomedical Engineering*,**39**(4), (2019) 967-982.
3. Xu J., Luo X., Wang G., Gilmore H., & Madabhushi A. A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images. *Neurocomputing*,**191**(2016), , 214-223.
4. Beevi K. S., Nair M. S., & Bindu G. R. Automatic mitosis detection in breast histopathology images using Convolutional Neural Network based deep transfer learning. *Biocybernetics and Biomedical Engineering*,*39*(1), (2019) 214-223.
5. Tashk A., Helfroush M. S., Danyali H., & Akbarzadeh-Jahromi M. Automatic detection of breast cancer mitotic cells based on the combination of textural, statistical and innovative mathematical features. *Applied Mathematical Modelling*, **39**(20), (2015) 6165-6182.
6. Öztürk Ş., & Akdemir B. HIC-net: A deep convolutional neural network model for classification of histopathological breast images. *Computers & Electrical Engineering*, **76**, (2019) 299-310.
7. Das D. K., & Dutta P. K. Efficient automated detection of mitotic cells from breast histological images using deep convolution neutral network with wavelet decomposed patches. *Computers in biology and medicine*,**104**, (2019) 29-42.
8. Öztürk Ş., & Akdemir B. Phase classification of mitotic events using selective dictionary learning for stem cell populations. *Computers & Electrical Engineering*,**67**, (2018) 25-37.
9. Kaushal C., Bhat S., Koundal D., & Singla A. Recent trends in computer assisted diagnosis (CAD) system for breast cancer diagnosis using histopathological images. *IRBM*, **40**(4), (2019) 211-227.
10. Tandon A., Singh N. N., Brave V. R., & Sreedhar G. Image analysis assisted study of mitotic figures in oral epithelial dysplasia and squamous cell carcinoma using differential stains. *Journal of oral biology and craniofacial research*,**6** (2016) S18-S23.

11. Bolhasani H., Amjadi E., Tabatabaeian M., & Jassbi S. J. A histopathological image dataset for grading breast invasive ductal carcinomas. *Informatics in Medicine Unlocked*,(2020). 100341.
12. Roy K., Banik D., Bhattacharjee D., & Nasipuri M. Patch-based system for Classification of Breast Histology images using deep learning. *Computerized Medical Imaging and Graphics*, **71**(2019). 90-103.
13. Gandomkar Z., Brennan P. C., & Mello-Thoms C. MuDeRN: Multi-category classification of breast histopathological image using deep residual networks. Artificial intelligence in medicine, **88** (2018). 14-24.
14. Roullier V., Lézoray O., Ta V. T., & Elmoataz A. Multi-resolution graph-based analysis of histopathological whole slide images: Application to mitotic cell extraction and visualization. *Computerized Medical Imaging and Graphics*, **35**(7-8), (2011) 603-615.
15. Maroof N., Khan A., Qureshi S. A., ul Rehman A., Khalil R. K., & Shim S. O. Mitosis detection in breast cancer histopathology images using hybrid feature space. Photodiagnosis and photodynamic therapy, **31** (2020) 101885.
16. Wahab N., Khan A., & Lee Y. S. Two-phase deep convolutional neural network for reducing class skewness in histopathological images based breast cancer detection. *Computers in biology and medicine*,**85** (2017) 86-97.

List of Publications

1. S. Balaji, T. Arunprasath, M. Pallikonda Rajasekaran, K. Sindhuja and R. Kottaimalai. “A Metaheuristic based Clustering Approach for Breast Cancer Identification for Earlier Diagnosis”, *International conference on smart electronics and Communication, IEEE*, (2023). 1445-1451 [Scopus Indexed].

2. S. Balaji, T. Arunprasath, M. Pallikonda Rajasekaran, G. Vishnuvarthanan and K. Sindhuja “Computer Aided Diagnostic System for Breast Cancer Detection Based on Optimized Segmentation Scheme and Supervised Algorithm” *Automatika* (2023) (Impact factor -1.790).

3. S. Balaji, T. Arunprasath, M. Pallikonda Rajasekaran and G.Vishnuvarthanan, “Supervised Learning Scheme based Mitotic Cells Detection and Classification in Breast Cancer Histopathological Images” *Soft Computing* (2023), (Accepted). (Impact factor - 3.732).